

經濟部 107 年度  
人工智慧產業關鍵技術拔尖計畫(1/4)  
合作研究計畫

中文任務導向型使用者對話模擬系統

建議書徵求文件

財團法人資訊工業策進會

中華民國 107 年 01 月 25 日

# 107年度合作研究計畫建議書徵求文件

## 一、 簡介

目前在AI的自然語理解技術大都以task-oriented對話系統進行研究，但若干障礙可能會阻礙研究和開發的進展。首先，研製的系統會在真實的場域環境與用戶進行互動，所以傳統的對話語料庫或文獻上用為benchmark的數據資料不能直接使用。其次，每個task都需要單獨的語料庫與特定的數據註釋。第三，task-oriented的對話收集和註釋需要廣泛的領域知識。因為構建一個合適的數據集既費錢又費時，一種常見的方法是基於一個例子對話的語料庫建立一個用戶模擬器 (User Simulation)。然後，當Dialogue Management與用戶模擬器交互時，可以最佳化Dialogue Policy；而後，經過訓練的Dialogue Management再跟真實使用者互動並進行適當的微調。

## 二、 計畫目標

透過本計畫研製一用戶模擬器(以供對話管理系統訓練使用)，針對一應用領域的中文對話情境，透過模擬器產生自然而合理的對話，而後才能與對話管理系統交互，進而擴展與最佳化調整既有對話管理的對話策略。

## 三、 計畫範圍

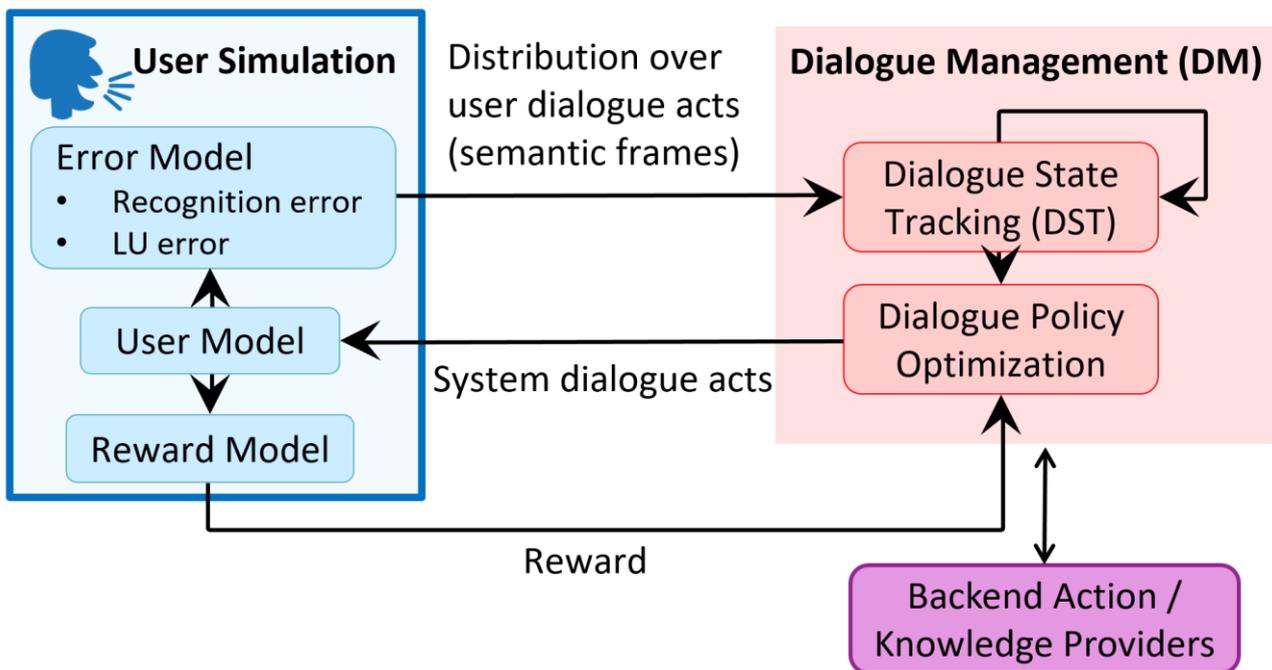


圖 1：系統流程圖

如圖1所示，整個互動包含模擬器與對話管理。模擬器包含錯誤處理模組(Error Model)、用戶模組(User Model)、獎勵模組(Reward Model)。對話管理包含對話狀態追蹤(Dialogue State Tracking)、對話策略最佳化(Dialogue Policy Optimization)，以及第三方的知識庫提供單元。模擬器與對話管理透過API方式相互傳遞訊令。此計畫目標為設計並訓練用戶模擬器，以供後續的對話系統做訓練用途。

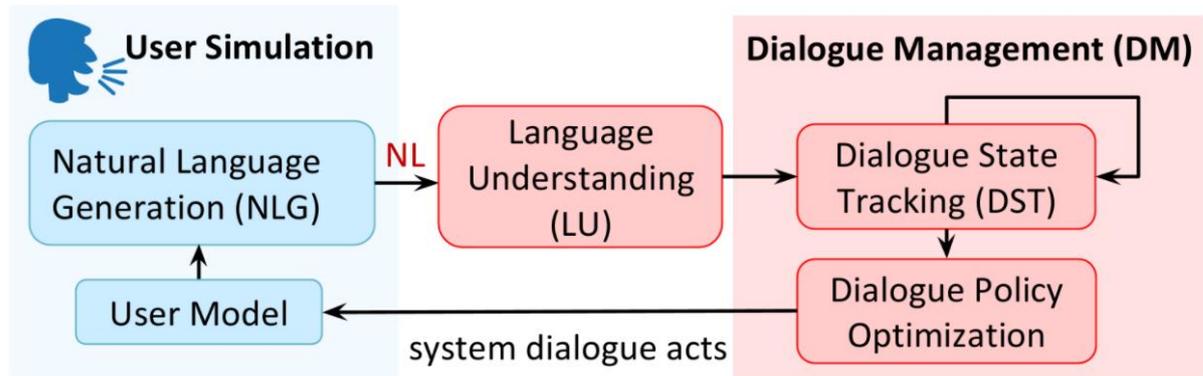


圖 2：加入語言生成模組之系統流程圖

更進一步，除了以語意的形式作互動以外，以自然語言方式互動才是系統主要的溝通模式，圖2為加入了語言生成（Language Generation）模組，並且讓系統端能夠先透過語言理解模組擷取語意後，再餵給對話管理進行互動之訓練及學習。

以下為此計畫主要的開發目標：

需求功能	功能說明
模擬器-錯誤模組	針對語意辨識錯誤與理解錯誤的處理
模擬器-用戶模組	根據 DM action 產生相對應的控制錯誤模組與控制獎勵模組
模擬器-獎勵模組	調整對話策略最佳化
模擬器-語言生成模組	根據用戶的動作產生相對應自然語言

#### ■系統開發環境

1. 作業系統：Linux Ubuntu X86\_64 作業系統
2. 資料庫：相容於MariaDB之關聯式資料庫 或 相容於MongoDB之非關聯式資料庫
3. 開發平台：Python Django Web Framework
4. 前端網頁需相容google chrome 63.xx或IE 11.xx
5. 技術架構：Python 3
6. 支援Nvidia Cuda 8.0 以上GPU運算加速技術
7. 人工智能開發框架：Tensorflow 1.2 以上
8. 研發團隊於107年10月31日前將系統上線至本會提供之雲端主機，並進行開發測試

#### ■系統開發介面

1. 全系統模組皆提供RESTful API介面，以GET/POST/PUT/DELETE的方式操作
2. 資料輸出入格式為JSON，資料內容為UTF-8編碼

#### ■系統部署與展示

1. 系統以docker技術進行封裝(系統將執行在docker環境上)
2. 系統將在本會提供的雲端主機上進行在線展示

## 四、 預期成果

透過本計畫研製一模擬器與對話管理，針對一應用領域的中文對話情境，透過模擬器產生自然而合理的對話，與既有對話管理交互並擴展與調整對話管理的對話策略。預期成果包含：

- 期中報告(含系統需求規格與設計規格書初版) 1篇。
- 期末報告(含系統需求規格、系統設計規格、系統安裝手冊、系統操作手冊、系統程式RESTful API使用文件、系統程式原始碼下載網址) 1篇。
- 期末交付核心技術及雛形系統(含binary code及source code，且需提供環境建置設定檔案Dockerfile)。

※前述成果如有專利構想或專利申請產出時，需注意專利申請之新穎性(novelty)。因凡經公開發表之研發成果，如擬申請專利，須於公開發表後6個月內完成，前述成果如是以論文方式公開發表，將無法取得大陸與歐盟等國之專利。

## 五、 執行方式

計畫執行方式如下：

- 每兩週定期召開研究討論會議。
- 期中進度報告。
- 期末總結報告與雛形系統展示。

## 六、 計畫期程及預估計畫總經費

計畫執行區間：107年03月01日至107年12月15日

總經費：600,000元

## 七、 驗收標準(含教育訓練)

交付項目	交付內容	交付型態	交付期限
期中報告	系統需求規格初版 系統設計規格初版	電子檔	107年8月31日
期末報告	系統需求規格 系統設計規格 系統安裝手冊 系統操作手冊 系統程式RESTful API使用文件 系統程式原始碼下載網址 (含binary code及source code，且需提供環境建置設定檔案Dockerfile)	電子檔	107年12月15日

## 八、 技術能力需求

具備自然語言理解、中文處理、對話系統、知識庫、深度學習等技術專長或研究之學術機構與人員。

附件1：契約書格式

1-1：計畫書格式

1-2：經費動支報表

1-3：成果報告撰寫須知

1-4：報告格式

1-5：論文格式

1-6：保密聲明書

1-7：委託匯款同意書